WHAT IS CLAIMED IS:

1. A document extracting device, comprising:

a similarity computing device to acquire a plurality of documents to be candidates for extraction and computing all degrees of similarity between the documents; and

a document extracting device to extract a combination of documents whose sum of the degrees of similarity between the documents computed by the similarity computing device is the smallest when any number of documents are extracted from among a group of the documents.

2. The document extracting device according to Claim 1,

the similarity computing device comprising:

a character-string-dividing functional unit to divide each of the documents into predetermined character strings;

a character-string frequency computing functional unit to compute document vectors of the documents on the basis of the frequency of appearance of the character strings divided by the character-string-dividing functional unit; and

a mutual similarity computing functional unit to compute the degrees of similarity between the documents on the basis of the document vectors obtained from the character-string frequency computing functional unit.

3. The document extracting device according to Claim 2,

the character-string-dividing functional unit dividing each of the documents into predetermined character strings using any one of character string division methods.

4. The document extracting device according to Claim 2,

the character-string frequency computing functional unit generating document vectors obtained by weighting each of the documents by TFIDF on the basis of the frequency of appearance of the divided character strings.

5. The document extracting device according to Claim 2,

the mutual similarity computing functional unit computing the degrees of similarity between the documents by a vector space method on the basis of the document vectors of the documents.

6. A document extracting program allowing a computer to serve as:

similarity computing device to acquire a plurality of documents to be candidates for extraction and computing all degrees of similarity between the documents; and

document extracting device to extract a combination of documents whose sum of the degrees of similarity between the documents computed by the similarity computing

device is the smallest when any number of documents are extracted from among a group of the documents.

7. The document extracting program according to Claim 6,

the similarity computing device comprising:

a character-string-dividing function to divide each of the documents into predetermined character strings;

a character-string frequency computing function to compute document vectors of the documents on the basis of the frequency of appearance of the character strings divided by the character-string-dividing function; and

a mutual similarity computing function to compute the degrees of similarity between the documents on the basis of the document vectors obtained by the character-string frequency computing function.

8. A document extracting program according to Claim 6,

the similarity computing device comprising:

a character-string-dividing function to divide each of the documents into character strings using any one of character string division methods;

a character-string frequency computing function to generate document vectors obtained by weighting each of the documents by TFIDF on the basis of the frequency of appearance of the divided character strings; and

a mutual similarity computing function to compute the degrees of similarity between the documents by a vector space method on the basis of the document vectors of the documents.

9. A document extracting method, comprising:

a plurality of documents to be candidates for extraction are acquired;

all degrees of similarity between the documents are computed; and

when any number of documents are extracted from among a group of the documents, a combination of documents whose sum of the degrees of similarity between the documents is the smallest is extracted.

10. The document extracting method according to Claim 9,

each of the documents being divided into predetermined character strings, the frequency of appearance of the divided character strings is computed, document vectors of the documents are computed on the basis of the frequency of appearance of the character strings, and then the degrees of similarity between the documents to be candidates for extraction are computed using the document vectors.

11. The document extracting method according Claim 9,

each of the documents being divided into predetermined character strings using any one of character string division methods, such as a morphological analysis method, an n-gram method, and a stop-word method, document vectors of the documents obtained by weighting each of the documents by TFIDF on the basis of the frequency of appearance of the divided character strings are computed, and the degrees of similarity between the documents to be candidates for extraction are computed using a vector space method on the basis of the document vectors.

12. A document extracting device, comprising:

a similarity computing device to acquire a plurality of documents to be candidates for extraction and computing all degrees of similarity between the documents; and

a document extracting device to extract a combination of documents based on the degrees of similarity between the documents computed by the similarity computing device when any number of documents are extracted from among a group of the documents.

13. A document extracting program allowing a computer to serve as:

similarity computing device to acquire a plurality of documents to be candidates for extraction and computing all degrees of similarity between the documents; and

document extracting device to extract a combination of documents based on the degrees of similarity between the documents computed by the similarity computing device when any number of documents are extracted from among a group of the documents.

14. A document extracting method, comprising:

a plurality of documents to be candidates for extraction are acquired;

all degrees of similarity between the documents are computed; and

when any number of documents are extracted from among a group of the documents, a combination of documents based on the degrees of similarity between the documents is extracted.